

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2004-080625

(43)Date of publication of application : 11.03.2004

(51)Int.Cl.

H04L 1/00  
G10L 11/02  
G10L 19/00  
H04M 3/00

(21)Application number : 2002-240821

(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD

(22)Date of filing : 21.08.2002

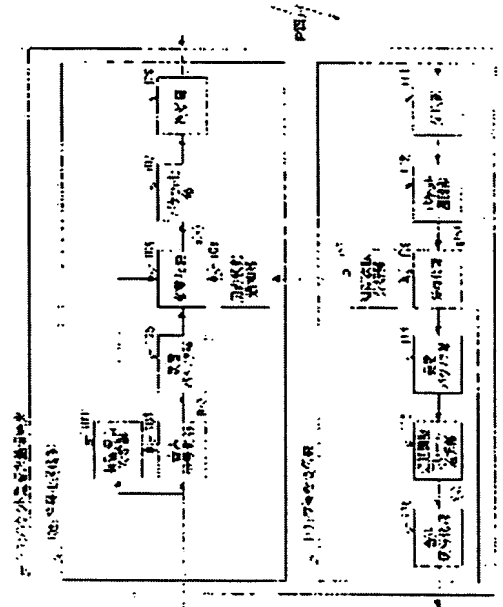
(72)Inventor : KAWASHIMA TAKUYA  
YOSHIDA KOJI

## (54) PACKET TYPE VOICE COMMUNICATION TERMINAL

### (57)Abstract:

PROBLEM TO BE SOLVED: To suppress deterioration in the voice quality and smoothly control the delay of a voice dynamically according to the circuit state.

SOLUTION: At a transmission side, a voice/silence decision section 103 analyzes an input voice and decides whether the input voice is voiced or voiceless. A voice encoding section 104 encodes the input voice conforming to the decided result of the voice/silence decision section 103. A multiplexing section 106 switches the multiplex number and the multiplexing depth confirming to the circuit state and the decided result of the voice/silence decision section 103 when selecting the encoded data to be multiplexed from the encoded data stored in a transmitting buffer section 105. On a reception side, the encoded data extracted from a packet received from an IP network is stored in a reception buffer section 114. A delay adjustment frame selection section 115 selects the encoded data giving an optimal delay by using the voice/silence information including in the encoded data stored in the reception buffer, and transfers the encoded data to a voice decoding section 116.



## LEGAL STATUS

[Date of request for examination]

04.07.2005

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

BEST AVAILABLE COPY

(19) 日本国特許庁(JP)

## (12) 公開特許公報(A)

(11) 特許出願公開番号

特開2004-80625

(P2004-80625A)

(43) 公開日 平成16年3月11日(2004.3.11)

(51) Int.Cl.<sup>7</sup>

F I

テーマコード(参考)

H04L 1/00

H04L 1/00

E

5K014

G10L 11/02

H04M 3/00

B

5K051

G10L 19/00

G10L 9/00

N

H04M 3/00

G10L 9/00

D

審査請求 未請求 請求項の数 3 O L (全 15 頁)

(21) 出願番号

特願2002-240821 (P2002-240821)

(22) 出願日

平成14年8月21日(2002.8.21)

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(74) 代理人 100105050

弁理士 鷺田 公一

(72) 発明者 河嶋 拓也

石川県金沢市西念一丁目1番3号 株式会社松下通信金沢研究所内

(72) 発明者 吉田 幸司

神奈川県横浜市港北区綱島東四丁目3番1号 松下通信工業株式会社内

Fターム(参考) 5K014 BA10 EA01 FA11

5K051 AA02 BB01 CC00 DD09 EE02

FF06 HH16 HH27 JJ13

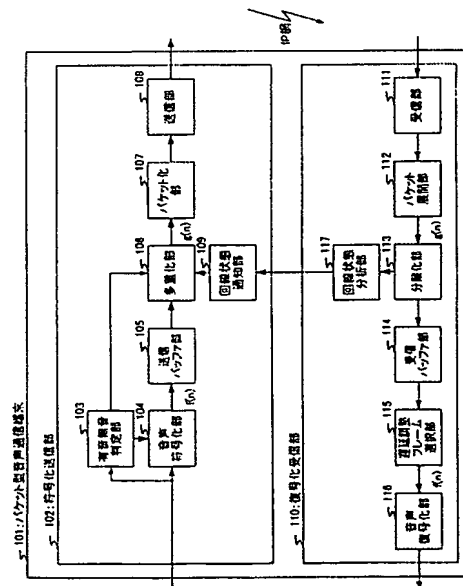
(54) 【発明の名称】 パケット型音声通信端末

(57) 【要約】

【課題】 回線状態に応じて、音質劣化を抑制しかつ音声の遅延をスムーズに動的制御すること。

【解決手段】 送信側では、有音無音判定部103にて、入力音声进行分析して有音か無音かが判定される。音声符号化部104では、入力音声を有音無音判定部103の判定結果に従って符号化する。多重化部106では、送信バッファ部105に蓄積された符号化データから多重化する符号化データを選択する際に、回線状態と有音無音判定部103の判定結果とに従って多重化数と多重化深度を切り替える。受信側では、IP網から受信されたパケットから取り出された符号化データが受信バッファ部114に蓄積される。遅延調整フレーム選択部115では、受信バッファに蓄積された符号化データに含まれている有音無音情報を利用して最適な遅延を与える符号化データを選択し、音声復号化部116に渡す。

【選択図】 図1



## 【特許請求の範囲】

## 【請求項 1】

入力音声进行分析して有音か無音かを判定する有音無音判定手段と、前記入力音声の前記有音無音判定手段の判定結果に従って符号化する音声符号化手段と、前記音声符号化手段が出力する符号化データを蓄積する送信バッファと、IP網に送出する多重化した符号化データを前記送信バッファに蓄積された符号化データを用いて生成する多重化手段であって、回線状態と前記有音無音判定手段の判定結果とに従って多重化数と多重化深度を切り替える多重化手段と、を具備することを特徴とするパケット型音声通信端末。

## 【請求項 2】

IP網から受信されたパケットから取り出された符号化データの多重化数と多重化深度が回線状態の他に有音時と無音時とで切替制御されている場合において、前記多重化数と多重化深度が切替制御されている符号化データを蓄積する受信バッファと、前記受信バッファに蓄積された符号化データから復号化する符号化データを選択するフレーム選択手段であって、前記受信バッファに蓄積された符号化データに含まれている有音無音情報を利用して最適な遅延を与える符号化データを選択するフレーム選択手段と、を具備することを特徴とするパケット型音声通信端末。

## 【請求項 3】

IP網から受信されたパケットから取り出された符号化データの多重化数と多重化深度が回線状態に応じて切替制御されている場合において、前記多重化数と多重化深度が切替制御されている符号化データを蓄積する受信バッファと、復号化する符号化データを選択するフレーム選択手段であって、運用遅延と連続フレーム消失数を使って最適な遅延を与える符号化データを選択するフレーム選択手段と、を具備することを特徴とするパケット型音声通信端末。

## 【発明の詳細な説明】

## 【0001】

## 【発明の属する技術分野】

本発明は、音声圧縮し、圧縮した符号化データをパケット化してインターネット網を伝送し、インターネット網から受信された符号化データを復号して音声通話を行うパケット型音声通信端末に関する。

## 【0002】

## 【従来の技術】

近年、インターネット技術の急速な発展／普及により、インターネットによるデータ伝送コストが急速に低下してきている。その一方で有線電話網は、通話品質（音質、安定性、低遅延）では勝るものの、高コスト性及び他サービスとの融合性の低さが問題となっている。そのため、電話サービスもインターネット上でサービスを行おうという機運が高まってきており、VoIP（Voice over Internet Protocol）の研究が盛んになってきている。既に音声等のリアルタイムサービス向けのプロトコル（RTP、RTCP、RSVP等）がIETF（The Internet Engineering Task Force）のRFC（Request for Comments）として規定されている。また、ITU-Tの規格としても、H.323という規格があり、徐々に普及してきている。

## 【0003】

ところが、インターネット網（以下「IP網」という）は、QoS（Quality of Service：サービス品質）が保証されないシステムであり、伝送パケットの到着時間の揺らぎや、伝送パケットの消失等の問題が頻繁に起こる。通常のデータであれば、パケットの到着時間の揺らぎは問題とはならない。その理由は、パケットの消失に関してもTCP（Transmission Control Protocol）やアプリケーションレベルでの再送制御を用いれば目的のデータを受信することができるからである。

## 【0004】

10

20

30

40

50

しかしながら、音声通話やテレビ電話等のサービスは、大幅な遅延が許されないサービスである。これらのサービスには、通常、再送制御は遅延が大きすぎるために用いることはない。これらのサービス実現に向けてIP網に対してQoSを確保する手法に対する取り組みがなされ、また現状のIP網を用いた場合のパケット消失対策として、FEC (Feed-forward Error Correction) 手法が研究されている。

#### 【 0 0 0 5 】

以下に、図7を参照して、FEC手法を用いた従来のVoIPについて簡単に説明する。なお、図7は、従来のパケット型音声通信端末の構成を示すブロック図である。図7に示す従来のパケット型音声通信端末701は、符号化送信部702と復号化受信部709とを備えている。

10

#### 【 0 0 0 6 】

符号化送信部702は、音声を圧縮符号化する音声符号化部703と、音声符号化部703にて符号化されたデータや正規の符号化データを受信できなかったときに補間に使用する補間用データを蓄積する送信バッファ部704と、回線状態に合わせて送信バッファ部704から送信する符号化データを選択し多重化する多重化部705と、多重化データをIPパケット化するパケット化部706と、パケット化部706にてパケット化されたデータをIP網に送信する送信部707と、復号化受信部709にて生成された回線品質を多重化部705に通知する回線状態通知部708とを備えている。

#### 【 0 0 0 7 】

復号化受信部709は、IP網からIPパケットを受信する受信部710と、受信部710にて受信されたIPパケットを展開するパケット展開部711と、パケット展開部711から多重化音声情報を受け取り、各フレーム毎に音声符号化データを分離する分離化部712と、分離化部712にて分離化された音声符号化データを蓄積する受信バッファ部713と、受信バッファ部713に蓄積された音声符号化データから復号に使用する音声符号化データを選択するフレーム選択部714と、フレーム選択部714にて選択された音声符号化データを復号する音声復号化部715と、受信されたIPパケットの連続性等を分離化部712にて分離化された音声符号化データに基づき確認等することによって回線品質を分析し送信側に通知する回線状態分析部716とを備えている。

20

#### 【 0 0 0 8 】

以上のように構成される従来のパケット型音声通信端末701の主な動作について説明する。符号化送信部702の音声符号化部703では、G. 726, G. 728, G. 729, AMRといった音声圧縮アルゴリズムを用いて圧縮を行い、符号化データ $f(n)$ を生成する。なお、 $f(n)$ は、時刻Nにおける第nフレームの符号化データを表している。この符号化データ $f(n)$ は、送信バッファ部704に蓄積される。

30

#### 【 0 0 0 9 】

送信バッファ部704では、このように生成された符号化データが、過去Mフレーム分蓄積されるとする。送信バッファ部704に蓄積される符号化データのうち、 $f(n)$ を除く過去の符号化データ $[f^2(n-1), f^3(n-2), \dots, f^M(n-M+1)]$ は、FECデータとして用いられる。

#### 【 0 0 1 0 】

つまり、次の動作ブロックである多重化部705では、ある時刻Nでは、処理中の符号化データ $f(n)$ と例えば1つ前の符号化データ $f(n-1)$ とが $g(n) = f(n) + f(n-1)$ と多重化され、次の時刻N+1では、処理中の符号化データ $f(n)$ と次の符号化データ $f(n+1)$ とが $g(n+1) = f(n+1) + f(n)$ と多重化される。送信側がこのように多重化することによって、受信側では、多重化された符号化データ $g(n)$ が受信できなかった時でも、次の符号化データ $g(n+1)$ が受信できれば、送信側での符号化データ $f(n)$ を得ることができるので、第nフレームを補間することなく再生することができる。

40

#### 【 0 0 1 1 】

ここで、送信バッファ部704及び受信バッファ部713に蓄積されFECデータとして

50

用いられる過去の符号化データは、音声符号化部 703 にて符号化されたデータそのものである必要はなく、伝送帯域を節約するため、例えば、さらに高圧縮した符号化データを用いたり、重要なデータだけにしたりすることができる。つまり、過去の符号化データは、単なるコピーでない可能性がある。

【 0012 】

そのため、図 7 では、現在処理中のフレーム（第  $n$  フレーム）の 1 つ前のフレームのデータは、 $f^2(n-1)$  と表している。また、現フレームを含めて  $M$  フレーム分を蓄積する場合は、一番古い符号化データは、 $f^M(n-M+1)$  と表している。

【 0013 】

過去の符号化データが単なるコピーでない場合は、当然、受信側では、受信された符号化データに対応した動作をすることが必要となる。但し、以後の説明においては、理解を容易にするため、過去の FEC 用の符号化データは、符号化データのコピーであるものとして説明する。

【 0014 】

さて、3GPP TS 26.235 では、 $f(n)$  と  $f(n-1)$  とで多重化する方法が示されている。しかしながら、この方法では、IP 網でのパケット消失状況が一定でなく、例えば 2 パケット連続で消失することが多いような場合には、対策効果が非常に薄い。

【 0015 】

そのため、例えば文献「A New Adaptive FEC Loss Control Algorithm for Voice Over IP Applications (Padhye C.; Christensen K. J.; Moreno W.; Performance, Computing, and Communication Conference, 2000. IPCCC, '00. Conference Proceedings of the IEEE International, 2000; Page(s): 307-313)」では、IP 網の状態に合わせて動的に FEC 用符号化データを多重化制御を行う方法が提案されている。この方法に従えば、IP 網に対する帯域負荷と音声品質に与える影響とのバランスを配慮したサービスが可能となる。

【 0016 】

すなわち、図 7 において、回線状態通知部 708 は、受信側の回線状態分析部 716 を通じて回線の状態を取得し、または、制御用コマンドを通じて IP 網から直接回線の状態を取得し、その取得した回線状態を多重化部 705 に通知する。多重化部 705 では、通知された回線状態に応じて多重化数と多重化深度（ここでは、何フレーム前のデータを多重化するかという意味で用いる）を動的に制御する。以下に、動的制御を行う際の一例を示す。

【 0017 】

(A) 連続パケット消失が多く、回線に帯域上余裕がない場合には、式 (1) のように、多重化深度を増加する。

$$g(n) = f(n) + f(n-1) \rightarrow g(n) = f(n) + f(n-2) \\ \dots (1)$$

【 0018 】

(B) 連続パケット消失は多いが、回線に帯域上余裕がある場合には、式 (2) のように、多重化数と多重化深度を共に増加する。

$$g(n) = f(n) + f(n-1) \rightarrow g(n) = f(n) + f(n-1) + f(n-2) \\ \dots (2)$$

【 0019 】

(C) 連続パケット消失からランダム性消失に変化し、回線の帯域上余裕がさらに低下した場合は、式 (3) に示すように、多重化数と多重化深度を共に減少する。

$$g(n) = f(n) + f(n-1) + f(n-2) \rightarrow g(n) = f(n) + f(n-1) \\ \dots (3)$$

【 0020 】

(D) パケット消失が殆ど発生しない場合は、式(4)に示すように、多重化数と多重化深度を共に減少する。

$$g(n) = f(n) + f(n-1) + f(n-2) \rightarrow g(n) = f(n) + f(n-1) \quad (4)$$

【0021】

【発明が解決しようとする課題】

しかしながら、従来のパケット型音声通信端末では、多重化数や多重化深度の動的制御はできるが、再生における遅延を制御することができない。つまり、そのシステムにおいて、多重化深度を最大Mとしたならば、常に受信側では最初の符号化データ  $f(n)$  を受け取ってから、その最初の符号化データ  $f(n)$  含むM個分のパケットを受信した後でなければ、符号化データ  $f(n)$  を復号することができず、遅延が固定されてしまう、つまり設計自由度が小さいという問題がある。

【0022】

図8を参照して説明する。なお、図8は、図7に示す従来のパケット型音声通信端末において実施される多重化数と多重化深度の動的制御を説明する図である。図8において、横軸は時間軸であり、縦軸は多重化するパケットを表しており、図8(1)の四角内の数字は、フレーム番号を表している。また、図8では、最大深度  $P=4$  とし、パケット番号  $p=0$  から  $p=6$  までが、多重化数=4、多重化深度=4となっている。パケット番号  $p=7$  から  $p=12$  までが、多重化数=2、多重化深度=2となっている。パケット番号  $p=13$  から  $p=20$  までが、多重化数=2、多重化深度=4となっている。

【0023】

パケット番号  $p=0$  から  $p=6$  までは、パケット番号  $p=3$  の時、フレーム番号 3, 2, 1, 0 の符号化データが  $g(3) = f(3) + f(2) + f(1) + f(0)$  というように多重化されて伝送される。パケット番号  $p=0$  から  $p=6$  までは、多重化数が4、多重化深度が4であるので、符号化データ  $f(3)$  を復号するためには、最後の符号化データ  $f(3)$  が受信されるパケット番号  $p=6$  まで待つ必要がある。

【0024】

次に、パケット番号  $p=7$  から  $p=12$  までは、多重化数が2、多重化深度が2である。符号化データ  $f(9)$  を復号するためには、本来最後の符号化データ  $f(9)$  が受信されるパケット番号  $p=10$  において復号が可能である。しかしながら、その場合にはそれ以前のフレームを廃棄しなければならない、不自然な再生音声となってしまう。そのため、最大深度  $P=4$  にしたがってパケット番号  $p=12$  で符号化データ  $f(9)$  を再生しなければならない。

【0025】

かりに、最大深度を無視し、送信してきた多重化深度に合わせて復号した場合には、多重化深度が大きくなると、今度はその差分の分だけフレーム補間が必要であるので、この場合も不自然な再生音声となってしまう。

【0026】

以上のことから、パケットの消失が少なく、回線状態が良好な場合であっても、遅延を少なくすると、回線の劣化に十分に対応できないので、回線が劣化するワーストケースを考慮して遅延を多めに取らなければならない。したがって、上記のように、設計上多重化する一番過去の符号化データによって遅延が決定されるという問題がある。

【0027】

本発明は、かかる点に鑑みてなされたものであり、回線状態に応じて動的に制御する多重化深度に合わせて、復号する音声の遅延をスムーズに制御することができるパケット型音声通信端末を提供することを目的とする。

【0028】

【課題を解決するための手段】

本発明のパケット型音声通信端末は、入力音声进行分析して有音か無音かを判定する有音無音判定手段と、前記入力音声を前記有音無音判定手段の判定結果に従って符号化する音声

10

20

30

40

50

符号化手段と、前記音声符号化手段が出力する符号化データを蓄積する送信バッファと、IP網に送出する多重化した符号化データを前記送信バッファに蓄積された符号化データを用いて生成する多重化手段であって、回線状態と前記有音無音判定手段の判定結果とに従って多重化数と多重化深度を切り替える多重化手段と、を具備する構成を採る。

【 0 0 2 9 】

この構成によれば、送信側において、回線状態に加えて、無音時や有音と無音の切替時に多重化数と多重化深度を変更することができる。

【 0 0 3 0 】

本発明のパケット型音声通信端末は、IP網から受信されたパケットから取り出された符号化データの多重化数と多重化深度が回線状態の他に有音時と無音時とで切替制御されている場合において、前記多重化数と多重化深度が切替制御されている符号化データを蓄積する受信バッファと、前記受信バッファに蓄積された符号化データから復号化する符号化データを選択するフレーム選択手段であって、前記受信バッファに蓄積された符号化データに含まれている有音無音情報を利用して最適な遅延を与える符号化データを選択するフレーム選択手段と、を具備する構成を採る。

【 0 0 3 1 】

この構成によれば、受信側において、無音期間に多重化数・多重化深度の変化を検知し、有音が始まる際に無音フレームの廃棄、補間を行い、送信側が指定する多重化深度に復号する音声の遅延を合わせることができる。

【 0 0 3 2 】

本発明のパケット型音声通信端末は、IP網から受信されたパケットから取り出された符号化データの多重化数と多重化深度が回線状態に応じて切替制御されている場合において、前記多重化数と多重化深度が切替制御されている符号化データを蓄積する受信バッファと、復号化する符号化データを選択するフレーム選択手段であって、運用遅延と連続フレーム消失数を使って最適な遅延を与える符号化データを選択するフレーム選択手段と、を具備する構成を採る。

【 0 0 3 3 】

この構成によれば、送信側において、任意のタイミングで多重化数・多重化深度の変更を行っている場合に、受信側では、多重化数・多重化深度の変更を検知し、現在の運用遅延と多重化深度の差以上連続でパケットの受信に失敗した場合に、補間フレームの廃棄・追加を行うことで多重化深度に合わせてスムーズに復号する音声の遅延を制御することができる。

【 0 0 3 4 】

【発明の実施の形態】

本発明の骨子は、回線状態に応じて動的に多重化数と多重化深度を変更制御する場合に、多重化深度に合わせて復号する音声の遅延を制御することにより、パケットの消失が少ないときは、極力遅延を減らすことで通話のインタラクティブ性を高め、また回線状態が悪くパケットが消失しやすい回線状態のときは、遅延を増やすというデメリットを受け入れることでパケット消失によるフレーム補間を回避して復号音声の劣化を抑え、通話内容を極力確実に伝えることができるようにすることである。

【 0 0 3 5 】

以下、本発明の実施の形態を図面を参照して詳細に説明する。

【 0 0 3 6 】

(実施の形態 1)

図 1 は、本発明の実施の形態 1 に係るパケット型音声通信端末の構成を示すブロック図である。図 1 に示すパケット型音声通信端末 101 は、符号化送信部 102 と復号化受信部 110 とを備えている。

【 0 0 3 7 】

符号化送信部 102 は、有音無音判定部 103 と、音声符号化部 104 と、送信バッファ部 105 と、多重化部 106 と、パケット化部 107 と、送信部 108 とを備えている。

復号化受信部 110 は、受信部 111 と、パケット展開部 112 と、分離化部 113 と、受信バッファ部 114 と、遅延調整フレーム選択部 115 と、音声復号化部 116 と回線状態分析部 117 とを備えている。

【 0038 】

まず、符号化送信部 102 の動作について説明する。マイクロホン等によって入力された音声信号は、A/D変換され、フレーム単位で有音無音判定部 103 と音声符号化部 104 とに入力される。

【 0039 】

有音無音判定部 103 では、例えば、LPC（線形予測係数）分析やピッチ分析、振幅の変化等を用いて入力されたフレームが有音フレームか、無音フレームであるかを判定を行い、その判定結果を音声符号化部 104 と多重化部 106 とに出力する。 10

【 0040 】

音声符号化部 104 では、入力されたフレームを、有音無音判定部 103 からの判定結果が無音フレームであれば無音用に符号化を行い、有音無音判定部 103 からの判定結果が有音フレームであれば有音用に符号化を行い、圧縮した符号化データ  $f(n)$  を送信バッファ部 105 に出力する。

【 0041 】

符号化データ  $f(n)$  は、送信バッファ部 105 に蓄積される。ここで、多重化深度を最大  $M$  とすると、送信バッファ部 105 には、 $f(n-M+1)$  までの符号化データが蓄積される。但し、前述したように、あるフレーム  $n$  の時に送信バッファ部 105 に蓄積されている過去の符号化データ  $f(n-1)$ 、 $f(n-2)$ 、 $\dots$   $f(n-M+1)$  は、符号化データの完全なコピーである必要はない。 20

【 0042 】

回線状態通知部 109 は、復号化受信部 110 から例えばパケット消失数等の回線状況を受け取ると、その回線状況を多重化部 106 に通知する。

【 0043 】

多重化部 106 は、回線状態通知部 109 から通知される IP 網の劣化具合に関する情報に基づき送信バッファ部 105 に蓄積されている、現フレームの符号化データ  $f(n)$  に対し、FEC用のデータとして過去の符号化データを選択して多重化した符号化データ  $g(n)$  を出力する処理を行う。その際に多重化情報も併せて例えばヘッダ情報としてパッキングする。 30

【 0044 】

ここで、前述したように、有音時に単に多重化数や多重化深度を変化させたのでは、伝送帯域の無駄や遅延を大きくさせてしまう。そこで、多重化部 106 では、有音無音判定部 103 からの判定結果に従い、フレームが無音フレームである時もしくは無音フレームから有音フレームになった時に多重化数と多重化深度を変更するようになっている。

【 0045 】

パケット化部 107 では、多重化部 106 にて多重化されたデータを例えば RTP (Real Time Protocol) にパケット化し、さらに UDP (User Datagram Protocol) / IP (Internet Protocol) に変換する。このように IP パケット化されたデータは、送信部 108 から IP 網に送信される。 40

【 0046 】

次に、復号化受信部 110 の動作について説明する。受信部 111 は、IP 網から関係する IP パケットを受信し、パケット展開部 112 に送る。パケット展開部 112 は、受信された IP パケットを展開して多重化された符号化データを取り出し、分離化部 113 に渡す。

【 0047 】

分離化部 113 は、パケット展開部 112 から受け取った多重化音声情報を各フレーム毎の符号化データに分離し、受信バッファ部 114 と回線状態分析部 117 とに渡す。なお、復号時間に間に合わないデータ等は、この分離化部 113 にて廃棄される。回線状態分 50



析部 117 は、例えば RTP を用いて消失パケット数等の回線状態を分析し、送信側の回線状態通知部 109 に渡す。

【 0048 】

受信バッファ部 114 では、分離化部 113 から受け取った符号化データが蓄積される。遅延調整フレーム選択部 115 は、例えば図 2 と図 3 に示す手順で、受信バッファ部 114 に蓄積された符号化データの中から有音フレームと無音フレームの情報を利用して最適な遅延調整フレームの符号化データを選択する。音声復号化部 116 は、遅延調整フレーム選択部 115 から受け取った符号化データを再生し、復号音声を出力する。

【 0049 】

図 2 と図 3 を参照して、遅延調整フレーム選択部 115 の動作を具体的に説明する。なお、図 2 は、多重化数と多重化深度がそれぞれ「4」から「2」に減少する場合の動作例を示し、図 3 は、多重化数と多重化深度がそれぞれ「2」から「4」に増加する場合の動作例を示している。

【 0050 】

まず、多重化深度が減少する場合の動作を説明する。図 2 において、図 2 (4) : パケット番号  $p$  は、「0」～「23」までが示されている。そのうち、パケット番号  $p = 0 \sim p = 9$  までが、多重化数および多重化深度が 4 のフレームであり、パケット番号  $p = 14 \sim p = 23$  までが、多重化数および多重化深度が 2 のフレームである。

【 0051 】

図 2 (1) : 送信側で生成される各符号化フレーム  $f(n)$  は、識別情報としてフレーム番号の他に、有音フレームであるか無音フレームであるかを示す有音無音情報を持っている。ここでは、符号化フレーム  $f(0)$  から  $f(6)$  までが有音フレームで、符号化フレーム  $f(7)$  から  $f(13)$  までが無音フレームで、符号化フレーム  $f(14)$  から  $f(23)$  までが有音フレームであるとしている。

【 0052 】

なお、無音フレームの区間に関しては、音声符号化方式により符号化データを送り続けるものや、無音区間を補間するのに十分な情報を間欠的に送り無音時に全く情報を送らないものもある。図 2 (1) では、無音情報を送り続けるようにしているが、もちろん間欠的に送るものでも構わない。

【 0053 】

図 2 (2) : 受信バッファ 114 には、多重化深度に応じた符号化データ  $g(n)$  が受信蓄積されることが示されている。すなわち、受信バッファ 114 には、パケット番号  $p = 0 \sim p = 9$  まで多重化深度 = 4 の符号化データ  $g(n)$  が受信蓄積され、パケット番号  $p = 10 \sim p = 13$  まで無音フレームのデータが格納され、パケット番号  $p = 14$  以降は、多重化深度 = 2 の符号化データ  $g(n)$  が受信蓄積される。また、無音時には、多重化していないことが示されている。勿論、有音時の多重化情報のまま多重化していても構わない。

【 0054 】

図 2 (3) : 遅延制御フレーム選択部 115 のフレーム選択動作を説明している。すなわち、パケット番号  $p = 0$  から  $p = 9$  までは、多重化深度が 4 であるので、最低でも 4 つのフレームを受信しなければ復号することができない。そのため、パケット番号  $p = 0, 1, 2$  では音声を復号せず、パケット番号  $p = 3$  で初めて符号化データ  $f(0)$  を全部受信できたため、その中から最も状態のいいものを選択して次の音声復号化部 116 に符号化データ  $f(0)$  を送ることができる。以降パケット番号  $p = 9$  で符号化データ  $f(6)$  が再生されるまでは同様に動作する。

【 0055 】

パケット番号  $p = 10$  から  $p = 13$  までの無音時では、受信した無音フレームを復号したり、もしくは、それ以前に受けたデータを元に補間動作を行う。パケット番号  $p = 14$  から  $p = 23$  までは、多重化深度が 4 から 2 に減少する。従来例であれば、多重化深度は最高値 4 に固定しなければならないため、符号化データ  $f(14)$  を復号するためには、パ

ケット番号  $p = 17$  まで待たなければならなかったが、今の例ではパケット番号  $p = 15$  で受信が完了しているので、パケット番号  $p = 15$  で復号が可能である。

【 0 0 5 6 】

これを実現するためには、従来例ではパケット番号  $p = 15$  で再生するはずだった符号化データ  $f(12)$  を復号せずに廃棄する必要がある。ところが、この符号化データ  $f(12)$  は、今の例では無音フレームであるので、復号せずに廃棄しても聴感上の劣化は無い。この例では、多重化深度が 4 から 2 へと変化したため、パケット番号  $p = 16$  のときに符号化データ  $f(13)$  も廃棄され、代わりに有音フレーム  $f(15)$  が選択される。以後そのままの遅延で復号されていく。

【 0 0 5 7 】

次に、多重化深度が増加する場合の動作を説明する。図 3 において、図 3 (4) : パケット番号  $p$  は、図 2 (4) と同様に「0」～「23」までが示されている。そのうち、パケット番号  $p = 0 \sim p = 7$  までが、多重化数および多重化深度が 2 のフレームであり、パケット番号  $p = 14 \sim p = 23$  までが、多重化数および多重化深度が 4 のフレームである。図 3 (1) : 符号化フレーム  $f(n)$  は、図 2 (1) と同内容である。

【 0 0 5 8 】

図 3 (2) : 受信バッファ 114 には、パケット番号  $p = 0 \sim p = 7$  まで多重化深度 = 2 の符号化データ  $g(n)$  が受信蓄積され、パケット番号  $p = 8 \sim p = 13$  まで無音フレームのデータが格納され、パケット番号  $p = 14$  以降多重化深度 = 4 の符号化データ  $g(n)$  が受信蓄積される。

【 0 0 5 9 】

図 3 (3) : 遅延調整フレーム選択部 115 のフレーム選択動作を説明している。すなわち、今度は、パケット番号  $p = 0 \sim p = 7$  までは、多重化深度が 2 であるので、パケット番号  $p = 0$  では復号されず、パケット番号  $p = 1$  で符号化データ  $f(0)$  が復号される。以後、パケット番号  $p = 7$  のときに符号化データ  $f(6)$  が復号されるまで同様である。パケット番号  $p = 8$  から  $p = 13$  までは、無音フレームであり、図 2 (3) と同様の動作を行う。

【 0 0 6 0 】

次のパケット番号  $p = 14$  で多重化深度が 2 から 4 に変化する。パケット番号  $p = 14$  では、変化以前の多重化深度 2 であれば、パケット番号  $p = 15$  で符号化データ  $f(14)$  が再生されるはずであるが、多重化深度が 4 であるため、この段階では後 2 フレーム待たねば符号化データ  $f(14)$  を受信することができない。

【 0 0 6 1 】

そのため、パケット番号  $p = 15$ 、16 では、無音フレームを補間することで多重化深度に遅延を合わせる。このように有音フレームが始まる前に無音フレームを補間してもほとんど劣化を感じることはないため、スムーズに運用遅延を変化させることができる。

【 0 0 6 2 】

以上のように、実施の形態 1 では、多重化深度を IP 網の状態に合わせて過去の符号化データを FEC 用に多重化して伝送するパケット型音声通信端末において、パケット消失対策として、符号化データを多重化して送信する場合に、有音無音情報を利用して多重化数、多重化深度を変更し、受信側で無音から有音へと変化する際に多重化方法に合わせて無音フレームの廃棄、補間を行うことによって復号する音声の遅延を切替制御できるようにしたので、異音を発生することなくスムーズに遅延を切り替えることができる。

【 0 0 6 3 】

これにより、パケット消失が少ない時は低遅延で、パケット消失が多い場合は、多重化深度を深くすることで遅延を増やして即時性を犠牲にしても確実に話の内容が伝わるようにするといった幅広い運用ができるようになる。

【 0 0 6 4 】

( 実施の形態 2 )

図 4 は、本発明の実施の形態 2 に係るパケット型音声通信端末の構成を示すブロック図で

10

20

30

40

50

ある。図 4 に示すパケット型音声通信端末 4 0 1 は、符号化送信部 4 0 2 と復号化受信部 4 0 9 とを備えている。

【 0 0 6 5 】

符号化送信部 4 0 2 は、音声符号化部 4 0 3 と、送信バッファ部 4 0 4 と、多重化部 4 0 5 と、パケット化部 4 0 6 と、送信部 4 0 7 と、回線状態通知部 4 0 8 とを備えている。復号化受信部 4 0 9 は、受信部 4 1 0 と、パケット展開部 4 1 1 と、分離化部 4 1 2 と、受信バッファ部 4 1 3 と、フレーム選択部 4 1 4 と、音声復号化部 4 1 5 と回線状態分析部 4 1 6 とを備えている。ここで、フレーム選択部 4 1 4 は、運用遅延記憶部 4 1 7 と、連続フレーム消失カウンタ部 4 1 8 と、遅延制御判定部 4 1 9 と、遅延調整フレーム選択部 4 2 0 とで構成されている。

10

【 0 0 6 6 】

まず、符号化送信部 4 0 2 の動作について説明する。マイクロホン等によって入力された音声信号は、A / D 変換され、フレーム単位で音声符号化部 4 0 3 に入力される。

【 0 0 6 7 】

音声符号化部 4 0 3 では、入力されたフレームを符号化し、圧縮した符号化データ  $f(n)$  を送信バッファ部 4 0 4 に出力する。符号化データ  $f(n)$  は、送信バッファ部 4 0 4 に蓄積される。ここで、多重化深度を最大  $M$  とすると、送信バッファ部 4 0 4 には、 $f(n - M + 1)$  までの符号化データが蓄積される。但し、前述したように、あるフレーム  $n$  の時に送信バッファ部 4 0 4 に蓄積されている過去の符号化データ  $f(n - 1)$ 、 $f(n - 2)$ 、 $\dots$   $f(n - M + 1)$  は、符号化データの完全なコピーである必要はない。

20

【 0 0 6 8 】

回線状態通知部 4 0 8 は、復号化受信部 4 0 9 から例えばパケット消失数等の回線状況を受け取ると、その回線状況を多重化部 4 0 5 に通知する。

【 0 0 6 9 】

多重化部 4 0 5 は、回線状態通知部 4 0 8 から通知される IP 網の劣化具合に関する情報に基づき送信バッファ部 4 0 4 に蓄積されている、現フレームの符号化データ  $f(n)$  に対し、FEC 用のデータとして過去の符号化データを選択して多重化した符号化データ  $g(n)$  を出力する処理を行う。その際に多重化情報も併せて例えばヘッダ情報としてパッキングする。

【 0 0 7 0 】

パケット化部 4 0 6 では、多重化部 4 0 5 にて多重化されたデータを例えば RTP (Real Time Protocol) にパケット化し、さらに UDP (User Datagram Protocol) / IP (Internet Protocol) に変換する。このように IP パケット化されたデータは、送信部 4 0 7 から IP 網に送信される。

30

【 0 0 7 1 】

次に、復号化受信部 4 0 9 の動作について説明する。受信部 4 1 0 は、IP 網から関係する IP パケットを受信し、パケット展開部 4 1 1 に送る。パケット展開部 4 1 1 は、受信された IP パケットを展開して多重化された符号化データ  $g(n)$  を取り出し、分離化部 4 1 2 に渡す。

【 0 0 7 2 】

分離化部 4 1 2 は、パケット展開部 4 1 1 から受け取った多重化音声情報を各フレーム毎の符号化データに分離し、受信バッファ部 4 1 3 と回線状態分析部 4 1 6 とに渡す。なお、復号時間に間に合わないデータ等は、この分離化部 4 1 2 にて廃棄される。回線状態分析部 4 1 6 は、例えば RTP を用いて消失パケット数等の回線状態を分析し、送信側の回線状態通知部 4 0 8 に渡す。

40

【 0 0 7 3 】

受信バッファ部 4 1 3 では、分離化部 4 1 2 から受け取った符号化データが蓄積される。フレーム選択部 4 1 4 は、運用遅延と連続フレーム消失数と使って遅延制御を行い、受信バッファ部 4 1 3 に蓄積された符号化データの中から最適な遅延調整フレームの符号化データを選択する。音声復号化部 4 1 5 は、フレーム選択部 4 1 4 から受け取った符号化デ

50

ータ  $f(n)$  を再生し、復号音声を出力する。

【 0 0 7 4 】

ここで、フレーム選択部 4 1 4 では、運用遅延記憶部 4 1 7 が、現在運用している遅延を記憶している。但し、この運用遅延は、送信側から送られてくる多重化深度とは必ずしも一致しない。連続フレーム消失カウンタ部 4 1 8 は、運用遅延と多重化深度が違う場合に機能し、受信フレームが連続で何フレーム消失したかをカウントする。このカウント値は、音声復号化部 4 1 5 において何フレーム連続フレーム消失補償するかと同値である。遅延制御判定部 4 1 9 は、運用遅延、受信フレームの多重化深度及び連続フレーム消失カウントを受取り、フレーム消失が連続で発生した時を利用してスムーズに運用遅延を変更できるように判定を行い、遅延調整フレーム選択部 4 2 0 に遅延制御の可否を伝える。遅延調整フレーム選択部 4 2 0 は、遅延制御を行うという判定を受けると、フレーム消失補償フレームの廃棄もしくは追加を行ったうえで、運用遅延を多重化深度にあわせるように動作する。

【 0 0 7 5 】

以下、図 5 と図 6 を参照して、フレーム選択部 4 1 4 の動作を具体的に説明する。なお、図 5 は、多重化数、多重化深度及び運用遅延がそれぞれ「4」から「2」に減少する場合の動作例を示し、図 6 は、多重化数、多重化深度及び運用遅延がそれぞれ「2」から「4」に増加する場合の動作例を示している。

【 0 0 7 6 】

まず、運用遅延が減少する場合の動作を説明する。図 5 において、図 5 ( 4 ) : パケット番号  $p$  は、「0」～「23」までが示されている。そのうち、パケット番号  $p = 0 \sim p = 7$  までが、多重化数、多重化深度及び運用遅延がそれぞれ「4」のフレームであり、パケット番号  $p = 8 \sim p = 23$  までが、多重化数、多重化深度及び運用遅延が「2」のフレームである。

【 0 0 7 7 】

図 5 ( 1 ) : 受信パケット  $g(n)$  の受信状態 ( 正常に受信できたか消失したかの状態 ) を示している。図 5 ( 1 ) では、受信パケット  $g(0)$  から  $g(9)$  までは正常に受信できたことを示している。受信パケット  $g(10)$  から  $g(13)$  まではフレーム消失による受信失敗を示している。受信パケット  $g(14)$  から  $g(23)$  までは正常に受信できたことを示している。

【 0 0 7 8 】

図 5 ( 2 ) : 受信バッファ部 4 1 3 には、多重化深度に応じた符号化データ  $g(n)$  が受信蓄積されることが示されている。すなわち、受信バッファ部 4 1 3 には、パケット番号  $p = 0 \sim p = 7$  まで多重化深度 = 4 の符号化データ  $g(n)$  が受信蓄積され、パケット番号  $p = 8$  以降多重化深度 = 2 の符号化データ  $g(n)$  が受信蓄積される。

【 0 0 7 9 】

図 5 ( 3 ) : フレーム選択部 4 1 4 のフレーム選択動作を説明している。すなわち、最初に受信したフレームの多重化深度で運用遅延を決めるとすると、運用遅延は 4 となる。従来例であればこのまま運用遅延は変更できない。今の例では、多重化深度が 4 から 2 に変更になった後に、パケット番号  $p = 10$  から  $p = 13$  のパケットを消失している。従来例であれば、符号化データ  $f(10)$  ,  $f(11)$  ,  $f(12)$  に相当するフレームについて音声復号化部 4 1 5 にてフレーム消失補償が行われ、パケット番号  $p = 16$  から符号化データ  $f(13)$  が運用遅延 4 のまま再生される。

【 0 0 8 0 】

それに対し、本発明によるフレーム選択部 4 1 4 では、次のようにして運用遅延の切り替えを行うようになっている。すなわち、符号化データ  $f(10)$  から  $f(12)$  は、受信できなかったためフレーム消失補償を行う。このとき、多重化深度は 2 であるので、パケット番号  $p = 14$  の段階で、符号化データ  $f(13)$  のデータを再生することが可能である。そして、パケット番号  $p = 14$  , 15 では、符号化データ  $f(11)$  ,  $f(12)$  に相当する補償フレームを廃棄することで、パケット番号  $p = 14$  で符号化データ  $f(13)$

）を復号することが可能となり、以後運用遅延を2に切り替えて運用することができる。  
この場合、少なくとも、パケット番号  $p = 13$  でフレーム消失補償が行われているので、  
運用遅延の変更が音質に大きな影響を与えない。

【0081】

但し、連続フレーム消失数が多重化深度の変化数よりも短いと、フレーム廃棄が行われる  
と、フレーム消失補償フレームが間に入らない状態となるので、復号音声は不自然になっ  
てしまう。また、間にフレーム消失補償フレームがあったとしても、ある程度以上の長さ  
の補償フレームがあった方が自然に聞こえる可能性がある。実運用にあたっては、シ  
ステムに合わせて遅延制御判定部419の判定アルゴリズムやパラメータを調整する必要  
がある。

10

【0082】

次に、運用遅延が増加する場合の動作を説明する。図6において、図6(4)：パケット  
番号  $p$  は、「0」～「23」までが示されている。そのうち、パケット番号  $p = 0 \sim p = 7$   
までが、多重化数、多重化深度及び運用遅延がそれぞれ「2」のフレームであり、パケ  
ット番号  $p = 8 \sim p = 23$  までが、多重化数、多重化深度及び運用遅延が「4」のフレ  
ームである。

【0083】

図6(1)：受信パケット  $g(n)$  の受信状態は、図5(1)と同様に、受信パケット  $g$   
(0) から  $g(9)$  までは正常に受信できたことを示している。受信パケット  $g(10)$   
から  $g(13)$  まではフレーム消失による受信失敗を示している。受信パケット  $g(14)$  20  
) から  $g(23)$  までは正常に受信できたことを示している。

【0084】

図6(2)：受信バッファ部413には、多重化深度に応じた符号化データ  $g(n)$  が受  
信蓄積されることが示されている。すなわち、受信バッファ部413には、パケット番号  
 $p = 0 \sim p = 7$  まで多重化深度 = 2 の符号化データ  $g(n)$  が受信蓄積され、パケット番  
号  $p = 8$  以降多重化深度 = 4 の符号化データ  $g(n)$  が受信蓄積される。

【0085】

図6(3)：フレーム選択部414のフレーム選択動作を説明している。すなわち、最初  
に受信したフレームの多重化深度で運用遅延を決めるとすると、運用遅延は2となる。従  
来例であればこのまま運用遅延は変更できないので、パケット番号  $p = 8$  以降では、多重 30  
化深度が増えたにもかかわらず、全ての符号化データが到着するのを待つことなく復号を  
開始している。今の例では、多重化深度が2から4に変更になった後に、パケット番号  $p$   
 $= 10$  から  $p = 13$  のパケットを消失している。従来例であれば、符号化データ  $f(10)$   
,  $f(11)$ ,  $f(12)$  に相当するフレームについて音声復号化部415にてフレ  
ーム消失補償が行われ、パケット番号  $p = 14$  から符号化データ  $f(13)$  が運用遅延2の  
まま再生される。

【0086】

それに対し、本発明によるフレーム選択部414では、次のようにして運用遅延の切り替  
えを行うようになっている。すなわち、符号化データ  $f(10)$  に関しては、完全に受信  
できなかったため、フレーム消失補償しなければならないが、符号化データ  $f(11)$ 、 40  
 $f(12)$  に関しては、パケット番号  $p = 14$  で受信できているため運用遅延を4に変更  
すれば復号が可能である。そこで、パケット番号  $p = 12$ 、13では、符号化データ  $f(11)$ 、  
 $f(12)$  に相当するフレーム消失補償を行いつつ、パケット番号  $p = 14$  から  
符号化データ  $f(11)$  を復号するようにしている。このようにすれば、スムーズに運用  
遅延を増やすことができる。

【0087】

以上のように、実施の形態2では、多重化深度をIP網の状態に合わせて過去の符号化デ  
ータをFEC用に多重化して伝送するパケット型音声通信端末において、パケット消失対  
策として、多重化された符号化データを受信する場合に、多重化深度の動的な変化に合  
わせて運用遅延を、連続フレーム消失期間を利用して切替制御するようにしたので、パケッ 50

ト消失が少ない時は低遅延で、パケット消失が多い場合は、多重化深度を深くすることで遅延を増やして即時性を犠牲にしても確実に話の内容が伝わるようにするといった幅広い運用ができるようになる。

【 0 0 8 8 】

【 発 明 の 効 果 】

以上説明したように、本発明によれば、IP網の状態に合わせて過去の符号化データをFEC用に多重化して伝送するパケット型音声通信端末において、有音無音情報や連続フレーム消失情報を使用して復号する音声の遅延をスムーズに切り替えることができる。

【 図 面 の 簡 単 な 説 明 】

【 図 1 】 本発明の実施の形態1に係るパケット型音声通信端末の構成を示すブロック図 10

【 図 2 】 図1に示す遅延調整フレーム選択部が有音フレームと無音フレームの情報を利用して最適な遅延調整フレームを選択する動作を説明する図（多重化深度が減少する場合）

【 図 3 】 図1に示す遅延調整フレーム選択部が有音フレームと無音フレームの情報を利用して最適な遅延調整フレームを選択する動作を説明する図（多重化深度が増加する場合）

【 図 4 】 本発明の実施の形態2に係るパケット型音声通信端末の構成を示すブロック図

【 図 5 】 図4に示すフレーム選択部が連続フレーム消失を利用して最適な遅延調整フレームを選択する動作を説明する図（運用遅延が減少する場合）

【 図 6 】 図4に示すフレーム選択部が連続フレーム消失を利用して最適な遅延調整フレームを選択する動作を説明する図（運用遅延が増加する場合）

【 図 7 】 従来のパケット型音声通信端末の構成を示すブロック図 20

【 図 8 】 図7に示す従来のパケット型音声通信端末において実施される多重化数と多重化深度の動的制御を説明する図

【 符 号 の 説 明 】

1 0 1、4 0 1 パケット型音声通信端末

1 0 2、4 0 2 符号化送信部

1 0 3 有音無音判定部

1 0 4、4 0 3 音声符号化部

1 0 5、4 0 4 送信バッファ部

1 0 6、4 0 5 多重化部

1 0 7、4 0 6 パケット化部 30

1 0 8、4 0 7 送信部

1 0 9、4 0 8 回線状態通知部

1 1 0、4 0 9 復号化受信部

1 1 1、4 1 0 受信部

1 1 2、4 1 1 パケット展開部

1 1 3、4 1 2 分離化部

1 1 4、4 1 3 受信バッファ部

1 1 5 遅延調整フレーム選択部

1 1 6、4 1 5 音声復号化部

1 1 7、4 1 6 回線状態分析部 40

4 1 4 フレーム選択部

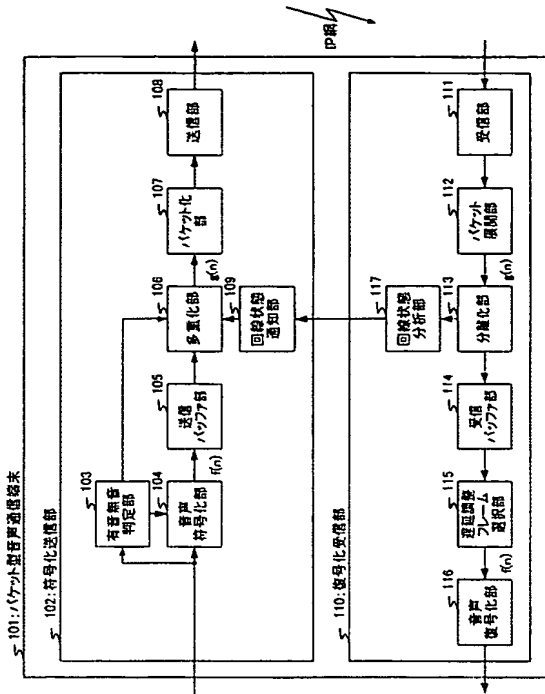
4 1 7 運用遅延記憶部

4 1 8 連続フレーム消失カウント部

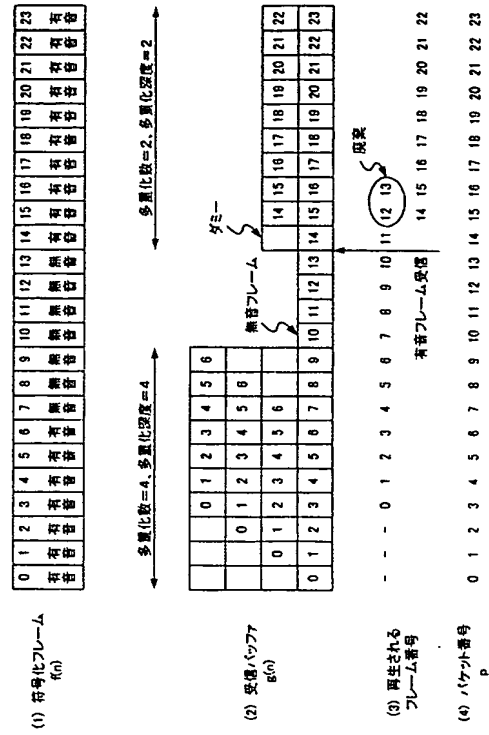
4 1 9 遅延制御判定部

4 2 0 遅延調整フレーム選択部

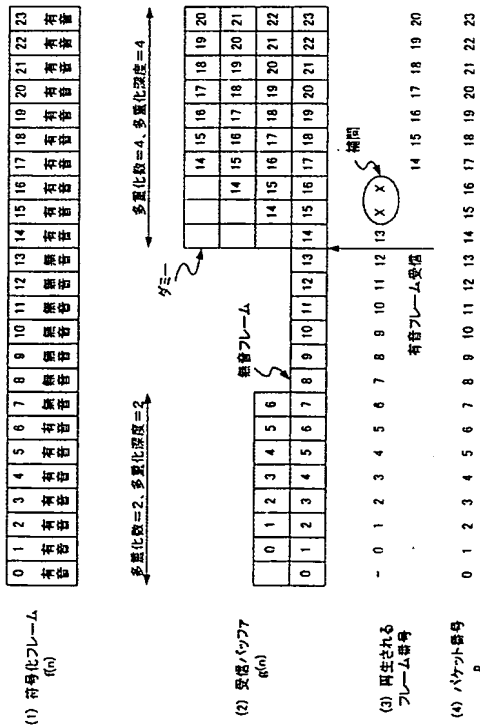
【 図 1 】



【 図 2 】



【 図 3 】



【 図 4 】

